A Structural Form for Higher-Index Semistate Equations. I. Theory and Applications to Circuit and Control Theory*

Kenneth D. Clark

Department of Computer Science
North Carolina State University
P.O. Box 8206
Raleigh, North Carolina 27695-8206

Submitted by N. K. Bose

ABSTRACT

We investigate the theoretical properties of a structural form for higher-index, linear time varying semistate systems

$$x' = A(t)x + B(t)u + f_1(t),$$

$$0 = C(t)x + D(t)u + f_2(t),$$

where $D(t)$ is singular but has constant rank for all $t \in [0, T]$. Many systems of interest in circuit and control theory exhibit the form discussed, in particular circuits with differentiators and linear-quadratic regulator problems with higher-order singular arcs. In addition to the semiexplicit structure, the systems we describe are characterized by a block upper Hessenberg coefficient $A(t)$, a zero pattern in $C(t)$ and $B(t)$, and an invertibility condition on the components of C, B, and the subdiagonal blocks of A. We show that there exists a coordinate system in which the Hessenberg form is equivalent to a semiexplicit index-1 system embedded between two linear chains of differentiators. In a companion paper, we establish the convergence and stability of backward differentiation formulas (BDF) applied to the index-4 Hessenberg form, extending the theory for the index-2 and index-3 special cases.

*Research supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-84-0240.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/88/$3.50

1. INTRODUCTION

In this paper, we consider a subclass of linear, time varying semiexplicit differential equations

$$x' = A(t)x + B(t)u + f_1(t), \tag{1.1a}$$

$$0 = C(t)x + D(t)u + f_2(t), \tag{1.1b}$$

where $D(t)$ is singular but has constant rank for all $t \in \Omega = [0, T]$, x is the *state* variable, u is the *algebraic* (or control) variable, and f_1, f_2 are known inputs. In the literature, systems of the form (1.1) are variously known as semistate, descriptor, differential-algebraic (DAE) or singular systems. Traditionally, they have arisen as models for circuits with linear dynamics and time varying components, in a singular-perturbation context as reduced-order models of multiple-time-scale systems, and in linear control theory either from system design or as necessary conditions for optimal control. More recently, applications of nonlinear semistate equations have appeared in trajectory control, fluid dynamics, and in the modeling of constrained mechanical systems [3, 27, 30].

Frequently, $D(t)$ in (1.1) is assumed to be invertible for all t of interest. In this case we say (1.1) is an *index*-1 system; otherwise it is *higher index* (we give precise definitions of the index of a singular system in the next section). Unlike ODEs, the solutions of higher-index systems involve derivatives of the inputs and coefficients, and the index is a direct measure of the number of differentiations present. Furthermore, some variables are determined by others via the constraint (1.1b) and its derivatives; hence not all initial values (x_0, u_0) admit smooth functional solutions. As such these systems pose difficulties not encountered in the numerical solution of ODEs [31, 32].

The most popular methods for numerically solving (1.1) and the more general *fully implicit* form

$$E(t)x' + F(t)x = f(t), \qquad t \in \Omega, \tag{1.2}$$

are the implicit *backward differentiation formulas* (BDF) [21]. If (1.2) is a solvable index-1 system, the BDF methods converge [24]. Furthermore, when E, F are constant, or if (1.1) is index 2 or 3 and the coefficients A, B, and C exhibit a specific structural pattern, the zero stable BDFs converge to the expected order of accuracy after an initial interval (or *discrete boundary* *layer*) of reduced-order convergence [4, 13, 27, 35]. The boundary layer is induced by inconsistent starting values, i.e., initial values which do not satisfy the algebraic constraints and their derivatives. The length of this interval is determined by the index and the order of the BDF method used. For example, the structures of the index-2 and index-3 systems analyzed in [4] and [27] are

$$\begin{aligned} x' &= A(t)x + B(t)u + f_1(t), \\ 0 &= C(t)x + f_2(t), \end{aligned} \tag{1.3}$$

where $C(t)B(t)$ is invertible (index 2), and

$$\begin{aligned} x' &= A_{11}(t)x + A_{12}(t)y + B_1(t)u + f_1(t), \\ y' &= A_{21}(t)x + A_{22}(t)y + f_2(t), \\ 0 &= C_2(t)y + f_3(t), \end{aligned} \tag{1.4}$$

where $C_2(t)A_{21}(t)B_1(t)$ is invertible (index 3). In (1.4) we have partitioned $B = (B_1^T, 0)^T$ and $C = (0, C_2)$. The kth-order BDF converges for (1.3) after $k+1$ steps and for (1.4) after $2k+1$ steps.

Unfortunately, BDFs do not converge for all higher-index systems, and moreover, necessary and sufficient conditions for convergence and stability do not exist. Attempts to derive such conditions have led to the synthesis of structural forms for singular systems which are solvable by BDFs. That is, rather than seek necessary and sufficient conditions for convergence, we attempt to classify systems according to their underlying structure. In turn, we expect such a classification will result in a greater understanding of these methods and, ultimately, the conditions we desire. This approach is twofold:

(1) identification of structural forms solvable by BDFs;
(2) identification of transformations which are BDF invariant.

We say a transformation T is a *BDF invariant* for the singular system (1.2) if the convergence and stability properties of BDF applied to (1.2) are preserved by T in the absence of rounding error. More precisely, let $x(\cdot)$ be a solution of (1.2), let $y(\cdot)$ be a solution after application of T, and let $\{x_n\}, \{y_n\}$ denote the corresponding BDF approximations. Then T is BDF invariant for (1.2) iff

$$\|x_n - x(t_n)\| = O(h^k) \quad \text{iff} \quad \|y_n - y(t_n)\| = O(h^k).$$

Notice that BDF invariance is defined relative to the pair (E, F), and also the length of the boundary layer is not affected by BDF invariant transformations. It follows from the definition of BDF [see equation (2.5)] that the *safe transformations* [5] of premultiplication by arbitrary (but smooth) invertible $P(t)$ and constant coordinate changes $x = Qy$ are BDF invariant for every pair (E, F), whereas in general, time varying coordinate changes or differentiations of (1.2) are not BDF invariants [9, 11].

The standard form for linear, constant-coefficient systems (1.2) is the Kronecker form of the regular pencil (E, F) [20]:

$$\begin{aligned} x_1' + Cx_1 &= f_1(t), \\ Nx_2' + x_2 &= f_2(t), \end{aligned} \tag{1.5}$$

where N is a direct sum of nilpotent Jordan blocks. The index of nilpotency of N is the index of (1.2). In this case, the existence of the form (1.5) (via constant safe transformations) is equivalent to solvability, but for solvable time varying systems $(E(t), F(t))$ need not be a regular matrix pencil for any t [7, Vol. 1, p. 140].

A time varying extension of (1.5) is developed in [5], [13] and can be written in compact form as

$$E(t)x_1' + F(t)x_1 = f_1(t), \tag{1.6a}$$

$$N(t)x_2' + x_2 = f_2(t) + G(t)x_1' + H(t)x_1, \tag{1.6b}$$

where (1.6a) is either an ODE [$E(t)$ nonsingular] or a solvable index-1 systems, and $N(t)$ is *structurally nilpotent*, i.e., $N(t)$ is strictly lower or upper triangular for all $t \in \Omega$. Note that (1.6) has the system-theoretical interpretation that the output x_1 of a generalized integrator is input to a generalized differentiator with output x_2, whereas (1.5) is a completely decoupled integrator-and-differentiator pair. Systems which are safely transformable to (1.6) can be solved by BDF methods. This includes all (theoretically) solvable index-2 systems in semiexplicit form [13]. However, not all systems of the form (1.4) are safely equivalent to (1.6) (see Section 3). Thus it is natural to consider whether (1.3), (1.4), and (1.6) are instances of the same form or are distinct in the sense there do not exist BDF invariant transformations which transform one form into another.

In this paper, we define and analyze a structural form for higher index semistate systems of the form (1.1) which includes (1.3) and (1.4) as special cases. These systems are characterized by a block upper Hessenberg coefficient $A(t)$, a zero pattern in $C(t)$ and $B(t)$, and an invertibility condition on the components of C, B and the subdiagonal blocks of A. Hence we refer to this structure as the *Hessenberg form*. Many systems of interest in circuit and control theory exhibit this structure, including circuits with operational amplifiers and linear-quadratic regulator problems with higher-order singular arcs. In addition to presenting a number of interesting theoretical properties, we show that the Hessenberg form and systems of the form (1.6) are distinct with respect to safe transformations. However, under a less restricted set of transformations (which are not BDF invariant for all semiexplicit systems), the Hessenberg form is equivalent to a coupled differentiator-integrator-differentiator triple. This suggests an expanded form consisting of a chain of coupled differentiators and integrators that would include (1.6) as a special case. Such a form is discussed in [9] and is currently the most promising form of higher-index systems solvable by BDFs. In a companion paper [15], we prove the convergence and stability of BDFs for the index-4 Hessenberg form, extending the theory for the index-2 and index-3 cases.

2. PRELIMINARIES

We refer the reader to [7] and the references therein for more detailed discussions of the review given in this section. In the sequel we assume that E, F, and f in (1.2) are real matrix- and vector-valued functions of the real parameter $t \in \Omega = [0,T]$ with dimensions $m \times m$ and $m \times 1$, respectively. The space of s times continuously differentiable functions on Ω is denoted by $C^s(\Omega)$. For notational convenience, we let $(x_1, x_2, \dots, x_r)^t = (x_1^T, x_2^T, \dots, x_r^T)^T$, where x_i is a column vector and $(\cdot)^T$ denotes transposition. Also, $R(E)$ and $N(E)$ are the range and nullspace, respectively, of the linear transformation E, while for two linear vector spaces A and B, $A \times B = \{(x_1, x_2)^t | x_1 \in A, x_2 \in B\}$ is the componentwise sum or Cartesian product space, $A \oplus B$ is the direct sum, and $A \ominus B$ is a complement of B in A.

By a solution of (1.2), we mean a function $x(\cdot)$ which is C^1 and satisfies (1.2) on an open subinterval $\vartheta \subset \Omega$. Then (1.2) is *solvable* iff for all sufficiently smooth f there exists at least one solution, all solutions are $C^1(\Omega)$, and each solution $x(\cdot)$ is uniquely determined by its value $x(t)$ for each $t \in \Omega$. In particular, solutions of (1.2) neither bifurcate nor escape to infinity in Ω.

The system (1.2) is *regular* in Ω iff the pencil (E, F) is regular, i.e., for each $t \in \Omega$ there exists a scalar $\lambda_t \in \mathbf{R}$ such that $\lambda_t E(t) + F(t)$ is invertible. Then the *local index* of (1.2) is the matrix index of $E_\lambda(t) = [\lambda_t E(t) + F(t)]^{-1}E(t)$, denoted by ind(1.2) or ind($E_\lambda(t)$). The index is independent of

λ_t [7]. If there is a fixed $\lambda \in \mathbf{R}$ such that $\lambda E(t) + F(t)$ is invertible for all $t \in \Omega$, we say (1.2) is *strictly* regular (s-regular). A sufficient (but not necessary) condition for s-regularity is that core-rank$(E_\lambda) = \operatorname{rank}(E_\lambda E_\lambda^D)$ is constant, where $(\cdot)^D$ denotes the *Drazin inverse* [11]. In this case, (1.2) is s-regular for all λ sufficiently large.

If (1.2) is a solvable index-1 system or an ODE, we say it is a *generalized integrator* (g-integrator) of f. It follows from [13] that (1.2) is a g-integrator of f if and only if $E(t)$ is singular or (1.2) is smoothly index 1, i.e., there exist smooth invertible $P(t)$ and $Q(t)$ such that

$$P(t)E(t)Q(t) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \qquad P(t)F(t)Q(t) = \begin{bmatrix} S(t) & 0 \\ 0 & I \end{bmatrix}. \tag{2.1}$$

Let $N(t)$ be structurally nilpotent for $t \in \Omega$. If $N(t) \neq 0$ for some t, then we say the system [equivalently, the pair (N, I)]

$$N(t)x' + x = f(t) \tag{2.2}$$

is a *generalized differentiator* (g-differentiator) of f. There are several ways to define the index of a differentiator.

DEFINITION 2.1. The *global index* σ_g of (2.2) is defined by

$$\sigma_g = \sup_{t \in \Omega} \operatorname{ind}(N(t)),$$

i.e., the global index is the maximum local index. The *structural index* of (2.2) is the smallest positive integer σ_s such that every σ_s-product

$$M_1(t)M_2(t) \cdots M_{\sigma_s}(t)$$

of matrices $M_i(\cdot)$ with the same nonzero structure as $N(\cdot)$ is identically zero.

See [18] for a slightly different definition of structural index for (1.2). Both global and structural indices are defined for every differentiator, although the structural index may not exist if $N(\cdot)$ is an arbitrary nilpotent matrix-valued function. Since $N \neq 0$, it follows that $\sigma_g, \sigma_s \geqslant 2$, so that g-differentiators and g-integrators are distinct. The global index of a g-differentiator is equivalent to the global index defined as the number of differentiations of the algebraic constraints required to reduce (1.2) to an ODE (see Algorithm 4.1 in [24]; also, see [34]).

In general, the local, global, and structural indices of higher-index singular systems (1.2) are different, although most often (and in particular for our purposes) they are equivalent. For example, if $N(\cdot)$ in (2.2) is smooth, there is a countable collection of open intervals $\{\vartheta_i\}$ (whose union is dense in Ω) such that on each ϑ_i the local and global indices are the same. However, consider (2.2) where

$$N(t) = \begin{bmatrix} 0 & & \\ M & 0 & \\ 0 & M & 0 \end{bmatrix},$$

$M \neq 0$, $M^2 \equiv 0$, and $MM' \neq 0$. Then $\sigma_g = 2$ while $\sigma_s = 3$.

Suppose (2.2) has structural index σ_s, and let $\mathscr{D}$ denote the derivative operator. Then for each $f \in C^{\sigma_s}$ the unique solution to (2.2) is

$$x(t) = \sum_{i=0}^{\sigma_s - 1} (-1)^i [N(t)\mathscr{D}]^i f(t). \tag{2.3}$$

The formula (2.3) follows from the fact that $N\mathscr{D}$ is a nilpotent operator of index σ_s on C^∞ with a Neumann series expansion (2.3) for $(N\mathscr{D} + I)^{-1}$. Note, however, that (2.3) is valid if $f \in C^r$, $r \geqslant \sigma_s$. Provided the coefficients are sufficiently smooth, if $f \in C^r$ is the input to a g-integrator then the solution will be C^r, while if f is input to a g-differentiator with structural index σ, then $x \in C^{r-\sigma+1}$, although the degree of differentiability will in general vary with the components.

In N in (1.6b) has the block form

$$N(t) = \begin{bmatrix} 0 & & & \\ N_{21} & 0 & & \\ \vdots & \ddots & \ddots & \\ N_{r-1,1} & \cdots & N_{r-1,r-2} & 0 \end{bmatrix}, \tag{2.4}$$

we say (1.6) is in *modified standard canonical form of size r* (MSCF-r). Then we have the following theorem.

THEOREM 2.1. *Let $E_\lambda = (\lambda E + F)^{-1}$ as before, and assume that $\operatorname{ind}(E_\lambda(t)) \leqslant r$ for $t \in \Omega$. Also let $N_0 = N(E_\lambda^r)$. Then the system* (1.2) *is safely transformable to MSCF-r iff the following conditions hold*:

(i) core-rank(E_λ) *is constant*;

(ii) *there exists a chain of constant subspaces $\{N_i\}$ for $i = 1, \ldots, r$ such that $N_r = \{0\}$, and for $i = 0, \ldots, r-1$ the N_i satisfy the containments*

$$N_i \supset N_{i+1} \quad \text{and} \quad E_\lambda N_i \subseteq N_{i+1}.$$

See [5] or [12] for a complete proof of Theorem 2.1. If (i) and (ii), hold let Q be the matrix whose columns are the basis vectors relative to the decomposition

$$(\mathbf{R}^m \ominus N_1) \oplus (N_1 \ominus N_2) \oplus \cdots \oplus (N_{r-2} \ominus N_{r-1}) \oplus N_{r-1}.$$

Then the safe transformations $x = Qy$ and premultiplication by $P(t) = Q^{-1}[\lambda E(t) + F(t)]^{-1}$ put (1.2) into the form (1.6), (2.4). Also, the number of nonzero subspaces in the chain and the fact that $\operatorname{ind}(E_\lambda) \leqslant r$ guarantee that the index of (1.6a) is $\leqslant 1$, while (i) assures that (1.6a) is solvable.

Finally, let $\{t_n\}$ be an equally spaced mesh on Ω with fixed steplength h. Then the k-step BDF applied to (1.2) is [35]

$$(E_n + h\beta_0 F_n)x_n = \sum_{i=1}^{k} \gamma_i E_n x_{n-i} + h\beta_0 f_n, \tag{2.5}$$

where x_n is the approximation to $x(t_n)$, and G_n denotes the exact evaluation $G(t_n)$ for $G = E, F$, and f. The coefficients β_0 and $\{\gamma_i\}$ depend on k and satisfy

$$\beta_0 = \left(\sum_{i=1}^{k} \frac{1}{i}\right)^{-1}, \qquad \gamma_i = (-1)^{i+1}\beta_0 \sum_{j=i}^{k} \left[\frac{1}{j}\binom{j}{i}\right], \qquad \sum_{i=1}^{k} \gamma_i = 1. \tag{2.6}$$

The following results will be useful in the sequel.

Proposition 2.1. *If $Q(t)$ is a smooth invertible matrix such that*

(i) $E(t)Q(t^*) = E(t)$ *for all* $t,\ t^* \in \Omega$,
(ii) $E(t)Q'(t) = 0$ *for all* $t \in \Omega$,

then the transformation $x = Q(t)y$ is a BDF invariant for (1.2).

Proof. Using (i), (ii) from above and Equation (2.5), the difference equation for y_n after the coordinate change $x = Qy$ is

$$(E_n + h\beta_0 F_n Q_n)y_n = \sum_{i=1}^{k} \gamma_i E_n y_{n-i} + h\beta_0 f_n. \tag{2.7}$$

But (i) implies $E_n Q_n^{-1} = E_n$ and $E_n Q_{n-i} = E_n$. Hence (2.7) is equivalent to

$$(E_n + h\beta_0 F_n)(Q_n y_n) = \sum_{i=1}^{k} \gamma_i E_n(Q_{n-i} y_{n-i}) + h\beta_0 f_n, \tag{2.8}$$

which, by the uniform boundedness of $\|Q(t)\|, \|Q(t)^{-1}\|$ on Ω, converges and is stable in $Q_n y_n$ iff (2.5) is convergent and stable in x_n. ■

COROLLARY. 2.1. *Every smooth invertible coordinate change $x = Q(t)y$ with Q of the form*

$$Q(t) = \begin{bmatrix} I & 0 \\ R(t) & S(t) \end{bmatrix} \tag{2.9}$$

is BDF invariant for the semiexplicit system (1.1). *The identity block in* (2.9) *is at least as large as x in* (1.1).

3. DESCRIPTION OF THE HESSENBERG FORM: PROPERTIES AND EXAMPLES.

In this section, we define the Hessenberg form and investigate some theoretical properties which prove useful for the numerical analysis in [15]. Several examples where this structure arises in applications are given, and it is shown how the form can be distinguished from other forms in the literature, in particular the MSCF-r defined in Section 2. Unless explicitly stated, for the remainder of this section we will assume $D(t) \equiv 0$ in (1.1) and refer to (1.1) in terms of the triple (A, B, C).

DEFINITION 3.1. The triple (A, B, C) in (1.1) is in the *Hessenberg form of size r* (denoted H_r) if $A(\cdot)$ is block upper Hessenberg,

$$A(t) = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1,r-2} & A_{1,r-1} \\ A_{21} & A_{22} & \cdots & A_{2,r-2} & A_{2,r-1} \\ & A_{32} & \cdots & A_{3,r-2} & A_{3,r-1} \\ & & \ddots & \vdots & \vdots \\ & & & A_{r-1,r-2} & A_{r-1,r-1} \end{bmatrix}, \tag{3.1a}$$

the matrices $B(t), C(t)$ have the form

$$B(t) = (A_{1r}(t), 0, \ldots, 0)^t, \quad C(t) = (0, \ldots, 0, A_{r,r-1}(t)), \quad (3.1b)$$

and the product Π_r defined by

$$\Pi_r = A_{r,r-1}A_{r-1,r-2} \cdots A_{21}A_{1r} \quad (3.2)$$

is invertible.

Partition the state variable $x = (x_1, x_2, \ldots, x_{r-1})^t$ according to the structure defined in (3.1a). We do not restrict the x_i to be the same size. That is, let $x_i \in \mathbf{R}^{n_i}$ for $i = 1, \ldots, r-1$ and $u \in \mathbf{R}^{n_r}$, so that (1.1) is a $p \times p$ system with $p = \sum n_i$. Note that the invertibility of Π_r implies $n_r = \min_j n_j$. Furthermore, each matrix R_j defined by

$$R_j = A_{r,r-1} \cdots A_{j+1,j}, \qquad j = 1, \ldots, r-1, \quad (3.3)$$

is $n_r \times n_j$ with full row rank n_r, while the matrices S_j defined by

$$S_j = \begin{cases} A_{j,j-1} \cdots A_{21}A_{1r}, & j = 2, \ldots, r-1, \\ A_{1r}, & j = 1, \end{cases} \quad (3.4)$$

satisfy the condition that S_j is $n_j \times n_r$ with full column rank n_r. It follows from the definition that $\Pi_r = R_j S_j$ for $j = 1, \ldots, r-1$.

THEOREM 3.1. *Let $A_{ij} \in C^r(\Omega)$. Then the system* (1.1) *with the structure* (3.1) *is solvable with local and global index r iff Π_r is invertible.*

The proof of Theorem 3.1 follows directly by differentiation and row reduction of (1.1) to an ODE system using Algorithm 4.1 in [24]. In fact, this reduction process terminates in r steps with an invertible coefficient of $(x', u')^t$ iff Π_r is invertible.

In the literature on control theory, (1.1) is said to be *totally singular* when $D(t) \equiv 0$. Using Theorem 3.1 and Dolezal's theorem [17], we can enlarge the applicability of the Hessenberg class of systems to include the case where $D = A_{rr} \neq 0$ but has constant rank by reducing the partially singular problem to a totally singular problem on a lower dimensional control space [16]. That is, there exist smooth invertible $P(t)$ and $Q(t)$ such that

$$P(t)D(t)Q(t)=\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.5}$$

It follows that there is a coordinate change of the form (2.9) which, followed by the row operation $P(t)$ on the algebraic equation (1.1b), puts (1.1) into the form

$$\begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x' \\ u_1' \\ u_2' \end{bmatrix}+\begin{bmatrix} A & B_1 & B_2 \\ C_1 & I & 0 \\ C_2 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ u_1 \\ u_2 \end{bmatrix}=\begin{bmatrix} f_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{3.6}$$

where A has the form (3.1a), $(C_1, C_2)^t=(0,\dots,0,PD)$, and $(B_1, B_2)=(A_{1r}Q,0,\dots,0)^t$. In addition to BDF invariance, these operations also preserve the index. Now eliminate B_1 and perform a symmetric permutation of the u_i to get

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x' \\ u_2' \end{bmatrix}+\begin{bmatrix} \tilde{A} & B_2 \\ C_2 & 0 \end{bmatrix}\begin{bmatrix} x \\ u_2 \end{bmatrix}=f, \tag{3.7a}$$

$$u_1=g_2-C_1x, \tag{3.7b}$$

where $\tilde{A}$ is block upper Hessenberg with the same subdiagonal blocks as A. Thus (3.7a) is a solvable system with index r and the triple $(\tilde{A}, B_2, C_2)$ has the structure (3.1). Let $C_2=(0,\dots,0,\underline{C}_2)$ and $B_2=(\underline{B}_2,0,\dots,0)^t$. From Theorem 3.1 we must have that $\underline{C}_2 A_{r-1,r-2}\cdots A_{21}\underline{B}_2$ is invertible; hence (3.7a) is an H_r system. Clearly, if x is computed to $O(h^j)$ accuracy by (2.5), then so is u_1. To summarize the previous discussion we state

THEOREM 3.2. *Let* (1.1) *be a partially singular problem, where $D \neq 0$ has constant rank in Ω, and assume (A, B, C) is an H_r triple. Then there exist BDF invariant transformations which put* (1.1) *into an H_r system with a control variable of strictly lower dimension, along with an algebraic constraint of the form* (3.7b).

EXAMPLE 3.1. Consider the linear-quadratic regulator problem

$$\min_{u\in U}\int_0^T\{\langle x(t),H(t)x(t)\rangle+\langle u(t),R(t)u(t)\rangle\}\,dt$$

subject to the conditions

$$x' = A(t)x + B(t)u, \quad x(0) = x_0, \qquad x(T) \text{ free or specified},$$

where U is a set of admissible controls (usually piecewise continuous functions, but for simplicity we assume continuity), $\langle \cdot, \cdot \rangle$ is the real inner product, and H and R are real symmetric, positive semidefinite in Ω. Assume $R(t)$ has constant rank in Ω. If (x, u) is optimal, then x, u satisfy the Euler-Lagrange equations

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x' \\ \lambda' \\ u' \end{bmatrix} = \begin{bmatrix} A(t) & 0 & B(t)^T \\ -H(t) & -A(t)^T & 0 \\ 0 & B(t)^T & R(t) \end{bmatrix} \begin{bmatrix} x \\ \lambda \\ u \end{bmatrix}, \tag{3.8}$$

$$x(0) = x_0, \qquad \lambda(T) = 0,$$

where λ is the costate (or Lagrange multiplier) variable. The order of the singular arc $(x(t), u(t))$ is p iff $u(t)$ appears explicitly after exactly $2p$ differentiations of (3.8) [1]. Equivalently, (3.9) is order p iff it has index $2p+1$ [8]. Thus (3.8) has a first-order singular arc iff $(\tilde{A}, B, B^T)$ is an H_3 triple, where

$$\tilde{A} = \begin{bmatrix} A & 0 \\ -H & -A^T \end{bmatrix},$$

i.e., iff $B^T H B$ is invertible. Higher-order singular problems with the Hessenberg structure can easily be constructed. For example, let

$$B = (B_1, 0)^t \quad \text{and} \quad H = \begin{bmatrix} 0 & 0 \\ 0 & H_{22} \end{bmatrix}.$$

Then (3.8) is safely equivalent to (coefficients only)

$$\begin{bmatrix} I & & & & \\ & I & & & \\ & & I & & \\ & & & I & \\ & & & & 0 \end{bmatrix}, \quad \begin{bmatrix} A_{11} & A_{12} & 0 & 0 & B_1 \\ A_{21} & A_{22} & 0 & 0 & 0 \\ & -H_{22} & -A_{22}^T & -A_{12}^T & 0 \\ & & -A_{21}^T & -A_{11}^T & 0 \\ & & & B_1^T & R \end{bmatrix}. \tag{3.9}$$

Also, (3.8), (3.9) has a 2nd-order singular arc iff $B_1^T A_{21}^T H_{22} A_{21} B_1$ is invertible. Finally, note that for any singular problem of order $p = (r-1)/2$ safely equivalent to the H_r form, the generalized Clebsch-Legendre sufficient conditions [16] are precisely

$$(-1)^r \Pi_r > 0. \tag{3.10}$$

EXAMPLE 3.2. Let A be an 8×8 block upper Hessenberg matrix of the form (3.1a) with $r = 4$ and

$$A_{21} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad A_{32} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad A_{ij} = 0 \quad \text{otherwise.}$$

Direct calculation shows that A is nilpotent index 3. Also, let $C = (0,\dots,0, [0\ 0\ 1])$ and $B = ([1\ 0\ 0], 0, \dots, 0)^t$. Then (A, B, C) is an H_4 triple. Clearly (A, B) is not completely controllable and (A, C) is not completely observable. On the other hand, if the control and output are scalar and (A, B) and (A, C) are simultaneously in controllability and observability canonical forms, then (A, B, C) forms a Hessenberg triple.

EXAMPLE 3.3. Consider the constrained nonlinear ODE $(r > 1)$

$$\begin{aligned} q^{(r-1)} &= f\big(q^{(r-2)}, q^{(r-3)}, \dots, q', q, t\big) + G(q, t)^T u, \\ 0 &= \Phi(q, t), \end{aligned} \tag{3.11}$$

where $G = \partial\Phi/\partial q$ and the number of constraints is no greater than the dimension of the coordinate variable q. If the constraints are linearly independent GG^T is invertible along trajectories. Written as a first-order system by letting $x_i = q^{(r-i-1)}$ for $i = 1, \dots, r-1$, (3.11) becomes

$$\begin{aligned} x' &= \begin{bmatrix} 0 & & & & \\ I & 0 & & & \\ & \ddots & \ddots & & \\ & & & I & 0 \end{bmatrix} x + \begin{bmatrix} f(x,t) + G(x_{r-1}, t)^T u \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\ 0 &= \Phi(x_{r-1}, t). \end{aligned} \tag{3.12}$$

If f and Φ are simultaneously linear in x and x_{r-1}, respectively, then (3.12) is in H_r and consequently has index r. Systems of the form (3.11) with $r = 3$ arise as force laws in the Euler-Lagrange formulation in Cartesian coordinates of constrained mechanical systems, e.g., rigid pendulums or robotic systems [27, 28]. The variable u can be interpreted as a Lagrange multiplier or as an unknown control parameter used to drive the state to the constraint manifold. Fourth-order ODEs (3.11) arise in the study of solid beams subject to bending forces [36]. In this case, q denotes the deflection of the beam from an equilibrium state, t is a spatial variable, and $\Phi(q,t) = 0$ is some desired shape. The resulting DAE (3.12) is index 5.

EXAMPLE 3.4 (Nonlinear Hessenberg forms). We can easily generalize the Hessenberg structure to nonlinear systems by letting Φ, $f = (f_1, f_2, \dots, f_{r-1})^t$ be nonlinear vector-valued functions of (x, u, t) and considering the system

$$\begin{aligned} x' &= f(x,u,t) \\ 0 &= \Phi(x_{r-1}, t). \end{aligned} \tag{3.13}$$

Then we say (3.13) is in nonlinear H_r form iff the linearization along trajectories

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x' \\ u' \end{bmatrix} = J(x(t), u(t), t) \begin{bmatrix} x \\ u \end{bmatrix} \tag{3.14}$$

where

$$J = \begin{bmatrix} \partial f/\partial x & \partial f/\partial u \\ \partial \Phi/\partial x & 0 \end{bmatrix},$$

has the structure (3.1) and the invertibility condition (3.2) on the Jacobians $A_{j,j-1} = \partial f_j/\partial x_{j-1}$ $(j = 2, \dots, r-1)$, $A_{1r} = \partial f_1/\partial x_r$, $A_{r,r-1} = \partial \Phi/\partial x_{r-1}$ is satisfied along trajectories. As in Theorem 3.2, Φ may depend on x_r provided $\partial \Phi/\partial x_r$ has constant rank.

Applications of the Hessenberg structure to problems in circuit theory and fluid dynamics can be found in [27] and will not be discussed here.

We noted in Section 1 that solvable semiexplicit index-2 systems (hence H_2) can be safely transformed into MSCF-2. Consider the special case of

(1.1) where (A, B, C) has the structure (3.1) and all state variables x_i have the same dimension. Interchanging rows 1 and r and permuting the variables according to $y_r = u$, $y_j = x_{r-j}$ $(j = 1, \ldots, r-1)$ puts (1.1) into the form

$$\begin{bmatrix} 0 & & & \\ I & 0 & & \\ & \ddots & \ddots & \\ & & I & 0 \end{bmatrix} \begin{bmatrix} y_1' \\ \vdots \\ y_r' \end{bmatrix} = \begin{bmatrix} A_{11} & & \\ \vdots & \ddots & \\ A_{r1} & \cdots & A_{rr} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_r \end{bmatrix} + f, \tag{3.15}$$

where the diagonal blocks A_{ii} are square and invertible. Thus (3.15) is safely equivalent to a structurally nilpotent system (2.2) with $\sigma_g = \sigma_s = r$ (i.e., MSCF-r). The next example shows that in general, the H_r and MSCF-r forms are distinct with respect to safe transformations.

EXAMPLE 3.5. Let (1.1) be the homogeneous system

$$\begin{aligned} x' &= y + B(t)u, \\ y' &= x, \\ 0 &= [1, t]y, \end{aligned} \tag{3.16}$$

where $(x, y) \in \mathbf{R}^2 \times \mathbf{R}^2$ is the state, $u \in \mathbf{R}$, and $B(t) = [1, t]^t$. This is an H_3 system with $\Pi_3 = (1 + t^2)$. Note that $R(B(t))$ varies with t and the coefficient of (x, y, u) in (3.16) is invertible for all t. Premultiply by its inverse to get

$$\begin{bmatrix} 0 & I & 0 \\ W & 0 & 0 \\ * & * & * \end{bmatrix} \begin{bmatrix} x' \\ y' \\ u' \end{bmatrix} = \begin{bmatrix} x \\ y \\ u \end{bmatrix}, \tag{3.17}$$

where $W = I - B(B^TB)^{-1}B^T$ is 2×2, $*$ is a generic entry whose value is unimportant, and

$$W(t) = \frac{1}{1+t^2} \begin{bmatrix} t^2 & -t \\ -t & 1 \end{bmatrix}.$$

Note that W is an orthogonal projector with varying nullspace. The subsystem

$$\begin{bmatrix} 0 & I \\ W & 0 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{3.18}$$

has index 2. Using Theorem 2.1 we will show that (3.18) is not safely transformable to MSCF-2 and hence (3.16) is not safely transformable to MSCF-3. Take $\lambda = 0$ and E_λ to be the coefficient in (3.18), so that

$$N_0 = N(E_\lambda^2) = \operatorname{span}\{(1,t,0,0)^t, (0,0,1,t)^t\}.$$

Hence $E_\lambda N_0 = \operatorname{span}\ \{(1,t,0,0)^t\}$ is time varying. Any minimal-dimensional subspace containing this space has dimension 2 and is not strictly contained in N_0. This violates condition (ii) of Theorem 2.1.

PROPOSITION 3.1. *Suppose* (1.1) *is a strictly regular, solvable index-3 system for* $t \in \Omega$ *and assume* $B(t)$ *has constant range. Then* (1.1) *is safely transformable to* MSCF-3.

Proof. Let $(E, F(t))$ be the pencil for (1.1), and premultiply (1.1) by $[\lambda E + F(t)]^{-1}$ to get

$$\begin{bmatrix} E_1(t) & 0 \\ E_2(t) & 0 \end{bmatrix}\begin{bmatrix} x' \\ u' \end{bmatrix} + \begin{bmatrix} I - \lambda E_1(t) & 0 \\ -\lambda E_2(t) & I \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \qquad \begin{matrix} (3.19\text{a}) \\ (3.19\text{b}) \end{matrix}$$

where

$$E_1(t) = (\lambda I - A)^{-1}\left\{I - B\left[C(\lambda I - A)^{-1}B\right]^{-1}C(\lambda I - A)^{-1}\right\}, \tag{3.20}$$

and the subsystem (3.19a) is solvable with index 2 [6]. But $N(E_1(t)) = R(B(t))$ so that (3.19a) is safely transformable to a semiexplicit index-2 system [13]. ■

COROLLARY 3.1. *If* $B(t) = (A_{13}, 0)^t$ *has constant range, then the* H_3 *triple* (A, B, C) *is safely transformable to* MSCF-3.

Proof. The system (1.1), (3.1) is strictly regular on every interval such that Π_r is invertible (see Theorem 3.3 below). The proof then follows from Proposition 3.1 for the case $r = 3$. ∎

The strict-regularity assumption in the previous results can be relaxed to pointwise regularity if we are only interested in transformability to MSCF-3 in a neighborhood of a given t. In terms of the regulator problem in Example 3.1, the above corollary and Theorem 3.2 imply all linear-quadratic regulators with first-order singular arcs are safely equivalent to (1.6), where (N, I) is an index-2 g-differentiator.

Properties of the Hessenberg Form

Variable-stepsize BDF strategies are generally unstable for higher-index systems unless the index is 2 [22, 31]. Hence strict regularity is an important property, since it guarantees that the coefficient matrix in (2.5) will be invertible for some fixed stepsize on the entire interval of integration. The next result shows that for any sufficiently small positive h, the matrix $E(t)+h\beta_0F(t)$ will be invertible for all $t \in \Omega$ if (E, F) has the Hessenberg form.

THEOREM 3.3. *Let (E, F) be the pencil for* (1.1), *where $D \equiv 0$ and (A, B, C) is an H_r triple* (3.1), (3.2). *Then (E, F) is strictly regular for $t \in \Omega$. Moreover, $(\lambda E + F)^{-1}$ is $O(\lambda^{r-1})$ as $\lambda \to \infty$.*

Proof. Let A, B, and C be given as in (3.1). Direct examination of the Schur complement (unreduced part) of $\lambda E + F$ after one block step of Gaussian elimination shows that (E, F) is strictly regular for λ sufficiently large if and only if $C(\lambda I - A)^{-1}B$ is. Take $\lambda > 1/\|A\|_\infty$, where $\|A\|_\infty = \sup_{t\in\Omega}\|A(t)\|$ for some matrix norm $\|\cdot\|$. Notice that by the structure of A, B, and C we have

$$CA^jB = 0,\ \ j = 0,\dots,r-3\quad (\text{for } r \geqslant 3)\quad \text{and}\quad CA^{r-2}B = \Pi_r\quad (\text{for } r \geqslant 2).$$

Hence, by expanding $(\lambda I - A)^{-1}$ in a power series we get that $C(\lambda I - A)^{-1}B = \lambda^{-(r-1)}(\Pi_r + G/\lambda)$, where $\|G\| = O(1)$. ∎

When (A, B, C) is constant, an immediate consequence of Theorem 3.3 is that the transfer-function matrix $T(s) = C(sI - A)^{-1}B$ of an H_r system has a zero at infinity of order $r-1$ in the frequency variable.

We now investigate some special algebraic properties of the solution set for H_r systems. Consider the case $r = 4$. After three steps of Algorithm 4.1 in

[24] we have the ODE system

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \tilde{f}_1 \\ f_2 \\ f_3 \end{bmatrix} \tag{3.21}$$

subject to

$$\begin{bmatrix} A_{43}A_{32}A_{21} & G_1 & G_2 \\ & A_{43}A_{32} & G_3 \\ & & A_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \tag{3.22}$$

$$u = \Pi_4^{-1}[g_4 - G_5 x_1 - G_6 x_2 - G_7 x_3], \tag{3.23}$$

where G_i, $\tilde{A}_{1j}$, g_k, and $\tilde{f}_1$ are matrix- and vector-valued functions of t which involve derivatives of the A_{ij}'s and f_i's and whose entries are smooth if the original coefficients are sufficiently smooth. Note that the first row of (3.21) is obtained from the first equation in (1.1), (3.1) by substituting (3.23). The actual values of these functions are unimportant for the subsequent analysis. The diagonal blocks of the constraint matrix in (3.22) are R_1, R_2, and R_3 as defined in (3.3). It follows that (3.22) is a $(3n_4)\times(n_1+n_2+n_3)$ linear system with full row rank. Since (3.21) is an ODE in $(x_1, x_2, x_3)^t$ and u is uniquely determined by (3.23), we have that the solution manifold for H_4 is solely determined by (3.22) and has rank ρ given by

$$\begin{aligned} \rho &= \dim N(R_1) + \dim N(R_2) + \dim N(R_3) \\ &= (n_1 - n_4) + (n_2 - n_4) + (n_3 - n_4). \end{aligned} \tag{3.24}$$

More generally, we have

PROPOSITION 3.2. *For $i = 1, \ldots, r-1$ define $\rho_i = \dim N(R_i) = n_i - n_r$, where R_i is defined in* (3.3). *Then the solution manifold for H_r has dimension $\rho = \Sigma \rho_i$.*

The proof is a direct extension of (3.24) by $r-1$ differentiations of the constraint and hence is omitted. The constraints analogous to (3.22), (3.23) are of the form

$$\begin{bmatrix} R_1 & * & \cdots & * \\ & R_2 & & * \\ & & \ddots & \vdots \\ & & & R_{r-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{r-1} \end{bmatrix} = g, \tag{3.25}$$

$$u = \Pi_r^{-1} w(x_1(t), \ldots, x_{r-1}(t), t), \tag{3.26}$$

where w is simultaneously linear in the x_i's. We should note that the arguments used to establish Proposition 3.2 hold for systems $x' = Ax + f$, $0 = Cx + g$, when the algebraic constraints are obtained by completely reducing the system (1.1) to an ODE by differentiation. However, for an arbitrary system $x' = Ax + f$, $0 = Cx + g$, the dimension of the solution manifold may be less than $\dim N(C)$.

The special structure of (3.1) leads to a natural direct-sum decomposition of the state variable x. From the definitions in (3.3), (3.4) we have

PROPOSITION 3.3. *For $i = 1, \ldots, r-1$ define $P_i = S_i \Pi_r^{-1} R_i$. Then P_i is the $n_i \times n_i$ rank-n_r projector onto $R(S_i)$ along $N(R_i)$. Thus $I - P_i$ is the complementary projector, and each space $\mathbf{R}^{n_i}$ admits the decomposition $\mathbf{R}^{n_i} = R(S_i) \oplus N(R_i)$.*

We can use Proposition 3.3 to compute consistent initial conditions on the constraint manifold by (3.25). Consider the case $r = 4$. For $i = 1,2,3$ let $x_i \in R(S_i)$ and thus $x_i = S_i q_i$ for some q_i. Since $R_i S_i = \Pi_4$, we can use back substitution in (3.22) to obtain a unique solution for the q_i's.

$$\begin{aligned} q_3 &= \Pi_4^{-1} g_3, \\ q_2 &= \Pi_4^{-1}\left[g_2 - G_3 S_3 \Pi_4^{-1} g_3\right], \\ q_1 &= \Pi_4^{-1}\left[g_1 - G_2 S_3 \Pi_4^{-1} g_3 - G_1 S_2 \Pi_4^{-1}\{g_2 - G_3 S_3 \Pi_4^{-1} g_3\}\right]. \end{aligned} \tag{3.27}$$

Hence an initial condition of (3.22) in $R(S_1) \times R(S_2) \times R(S_3)$ is $(S_1 q_1, S_2 q_2, S_3 q_3)^t$ evaluated at $t = 0$. This does not say that the component of $(x_1, x_2, x_3)^t$ in this subspace is uniquely determined, since in general

$$N(S) \neq N(R_1) \times N(R_2) \times N(R_3),$$

where S is the coefficient matrix in (3.22).

We conclude this section with an interesting characterization of the solution manifold for H_r systems which only involves (time varying) row operations and coordinate changes, and does not require differentiation of the algebraic constraint as in (3.25). This approach will also give some insight into why BDF methods should work for these systems.

First, observe that block upper triangular coordinate changes on the state variable preserve the Hessenberg structure. That is, let $Q(t)$ be defined by

$$Q(t)=\begin{bmatrix} Q_{11} & Q_{12} & \cdot & \cdot & \cdot & Q_{1,r-1} \\ & & \ddots & & & \vdots \\ & & & & & Q_{r-1,r-1} \end{bmatrix}, \tag{3.28}$$

where the diagonal blocks are invertible, and let $z=Qy$ in (1.1), where $z=(x,u)^t$. Then the resulting pencil after row reduction to semiexplicit form is $(E,\underline{F})$ where $\underline{F}$ has the form

$$\begin{bmatrix} * & * & \cdots & * & Q_{11}^{-1}A_{1r} \\ Q_{22}^{-1}A_{21}Q_{11} & * & \cdots & * & 0 \\ & Q_{33}^{-1}A_{32}Q_{22} & \cdots & * & 0 \\ & & \ddots & \vdots & \vdots \\ & & & A_{r,r-1}Q_{r-1,r-1} & 0 \end{bmatrix}. \tag{3.29}$$

In particular, block diagonal coordinate changes preserve the Hessenberg structure, although in general these transformations are not BDF invariant for all semiexplicit systems [11].

Assume $r>3$. Since $A_{r,r-1}$ and A_{1r} have full row and column rank n_r, respectively, there exist nonsingular $Q_{11}, Q_{r-1,r-1}$ such that

$$A_{r,r-1}Q_{r-1,r-1}=[I \quad 0], \qquad Q_{11}^{-1}A_{1r}=[I \quad 0]^t, \tag{3.30}$$

where the identity blocks are $n_r\times n_r$. Let $Q=\operatorname{diag}(Q_{ii})$, where Q_{11} and $Q_{r-1,r-1}$ are given as in (3.30), $Q_{ii}=I$ $(n_i\times n_i)$ otherwise. Setting $z=Qy$

and performing row operations gives the coefficient (of y only)

$$\left[\begin{array}{ccccc}
\begin{matrix} A^1_{11} & A^2_{11} \\ A^3_{11} & A^4_{11} \end{matrix} & \cdot & \cdot\ \cdot & \begin{matrix} A^1_{1,r-1} & A^2_{1,r-1} \\ A^3_{1,r-1} & A^4_{1,r-1} \end{matrix} & \begin{bmatrix} I \\ 0 \end{bmatrix} \\
\begin{bmatrix} A^1_{21} & A^2_{21} \\ A^3_{21} & A^4_{21} \end{bmatrix} & & & \vdots & \vdots \\
& \ddots & & & \\
& & \begin{bmatrix} A^1_{r-1,r-2} & A^2_{r-1,r-2} \\ A^3_{r-1,r-2} & A^4_{r-1,r-2} \end{bmatrix} & \begin{matrix} A^1_{r-1,r-1} & A^2_{r-1,r-1} \\ A^3_{r-1,r-1} & A^4_{r-1,r-1} \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\
& & & [I \quad 0] & 0
\end{array}\right], \tag{3.31}$$

where we have partitioned the entries according to (3.30) and the dimensions of the x_i. The coefficient of y' is the semiexplicit projector in (1.1). From Corollary 2.1, the left-to-right column operations which zero all entries in the first row of (3.31) but the last one are BDF invariant, as are the row operations which zero the third column from the right. The coefficient of y' is unchanged by these transformations. The invertibility condition on Π_r implies

$$\left[A^1_{r-1,r-2} \quad A^2_{r-1,r-2}\right] \cdot \tilde{A}_{r-2,r-3} \cdots \tilde{A}_{32} \cdot \left[A^1_{21} \quad A^3_{21}\right]^t \tag{3.32}$$

is invertible, where

$$\tilde{A}_{i,i-1} = \begin{bmatrix} A^1_{i,i-1} & A^2_{i,i-1} \\ A^3_{i,i-1} & A^4_{i,i-1} \end{bmatrix}.$$

Note also that the matrix in (3.32) is $n_r \times n_r$ and has two less factors than Π_r. Performing the above operations along with the column permutations (col. 1) $\to$ (col. $r-1$), (col. $r-2$) $\to$ (col. r), (col. $r-1$) $\to$ (col. $r-2$), (col. r) $\to$ (col. 1), and the interchange of rows $r-2$ and $r-1$, yields the coefficients

$$
\left[\begin{array}{c|cccccccc}
0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & I & 0 \\
\hline
0 & I & & & & & & 0 & 0 & 0 \\
0 & & I & 0 & & & & \cdot & \cdot & \cdot \\
\cdot & & 0 & I & & & & \cdot & \cdot & \cdot \\
\cdot & & & & \ddots & & & \cdot & \cdot & \cdot \\
\cdot & & & & & \ddots & & \cdot & \cdot & \cdot \\
\cdot & & & & & & \ddots & \cdot & \cdot & \cdot \\
\cdot & & & & & & & I & 0 & 0 \\
\cdot & & & & & & & 0 & 0 & I \\
\hline
0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0
\end{array}\right] \tag{3.33}
$$

and

$$
\left[\begin{array}{cccccccccccc}
I & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 \\
0 & A^4_{11} & A^3_{12} & A^4_{12} & \cdot & \cdot & \cdot & \cdot & \cdot & A^4_{1,r\text{-}1} & A^3_{11} & 0 \\
0 & A^2_{21} & A^1_{22} & A^2_{22} & \cdot & \cdot & \cdot & \cdot & \cdot & A^2_{2,r-1} & A^1_{21} & 0 \\
0 & A^4_{21} & A^3_{22} & A^4_{22} & \cdot & \cdot & \cdot & \cdot & \cdot & A^4_{2,r-1} & A^3_{21} & 0 \\
0 & 0 & A^1_{32} & A^2_{32} & \cdot & \cdot & \cdot & \cdot & \cdot & & 0 & 0 \\
0 & 0 & A^3_{32} & A^4_{32} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & & & & \ddots & & & & & \cdot & \cdot & \cdot \\
\cdot & & & & & *_1 & *_2 & & & \cdot & \cdot & \cdot \\
\cdot & & & & & *_3 & *_4 & & & \cdot & \cdot & \cdot \\
\cdot & & & & & 0 & 0 & A^3_{r-1,r-2} & A^4_{r-1,r-2} & A^4_{r-1,r-1} & 0 & 0 \\
\cdot & & & & & & & A^1_{r-1,r-2} & A^2_{r-1,r-2} & A^2_{r-1,r-1} & 0 & 0 \\
0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & 0 & I
\end{array}\right] \tag{3.34}
$$

where $*_i = A^i_{r-2,r-3}$. In the case $r = 5$, the subdiagonal block of the enclosed matrix in (3.34) is

$$
\begin{bmatrix} 0 & A^1_{32} & A^2_{32} \\ 0 & A^3_{32} & A^4_{32} \\ 0 & 0 & 0 \end{bmatrix},
$$

whereas in the case $r = 4$ this block is not present and we have that the enclosed block is simply

$$\left[\begin{array}{cccc|c} A_{11}^4 & A_{12}^3 & A_{12}^4 & A_{13}^4 & A_{11}^3 \\ A_{21}^2 & A_{22}^1 & A_{22}^2 & A_{23}^2 & A_{21}^1 \\ A_{21}^4 & A_{22}^3 & A_{22}^4 & A_{23}^4 & A_{21}^3 \\ 0 & A_{32}^3 & A_{32}^4 & A_{33}^4 & 0 \\ \hline 0 & A_{32}^1 & A_{32}^2 & A_{33}^2 & 0 \end{array}\right].$$

In any case the enclosed system has the Hessenberg structure of size $r-2$ and satisfies the requisite invertibility condition, hence it has index $r-2$. Note also that the dimension of the state variable x has been reduced by $2n_r$, i.e., the circled identity blocks in (3.33), (3.34) are $n_r \times n_r$. In the degenerate cases $r = 2,3$, the reduction yields the systems

$$\begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & 0 \\ \bar{A}_{21} & \bar{A}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (r = 2), \qquad (3.35)$$

$$\begin{bmatrix} 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & \bar{A}_{11} & \bar{A}_{12} & 0 \\ 0 & \bar{A}_{21} & \bar{A}_{22} & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \qquad (r = 3), \qquad (3.36)$$

where $\bar{A}_{22}$ is invertible. Clearly the enclosed systems in (3.35) and (3.36) are semiexplicit index 1. Finally, we observe that since the column corresponding to the control variable moves to the first column during the reduction step, in the ensuing step the identity block generated in the first row [see (3.33)] will move to the (1,2) position in the derivative coefficient. By induction we have just proved:

THEOREM 3.4. *Let* (1.1) *be in the* H_r *form, and define* $p = \langle (r-1)/2 \rangle$, *where* $\langle \cdot \rangle$ *is the greatest-integer function. Then there exist an integer* q *and sequences*

$$\{Z_i(t)\}_1^q, \qquad \{P_i\}_1^q, \quad \textit{and} \quad \{Q_i\}_1^q$$

of invertible matrices such that P_i *is a permutation;* Q_i *is block lower* *triangular of the form* $Q_i = \operatorname{diag}(I, \tilde{Q}_i, I)$, *where*

$$\tilde{Q}_i = \begin{bmatrix} Q_{11} & & & & & \\ \cdot & I & & & & \\ \cdot & & \ddots & & & \\ \cdot & & & I & & \\ \cdot & & & & Q_{k_i-1,k_i-1} & \\ Q_{k_i,l} & \cdot & \cdot & \cdot & Q_{k_i,k_i-1} & I \end{bmatrix};$$

and the transformations $z = Qy$ *and premultiplication by* $Z(t)$, *where*

$$Q = \prod_{i=1}^{q} (Q_i P_i), \qquad Z(t) = \prod_{i=q}^{1} Z_i(t),$$

put (1.1) *into the form*

$$\begin{bmatrix} 0 & I & 0 & \cdot & \cdots & \cdot & \cdot & 0 & * & \cdots & * \\ 0 & 0 & I & 0 & \cdots & \cdot & \cdot & 0 & * & \cdots & * \\ & & \ddots & \ddots & \ddots & & & \vdots & \vdots & & \vdots \\ & & & 0 & I & 0 & 0 & 0 & * & \cdots & * \\ & & & & 0 & 0 & I & 0 & * & & * \\ & & & & 0 & \boxed{I} & \boxed{0} & 0 & * & & * \\ & & & & 0 & \boxed{0} & \boxed{0} & I & * & \cdots & * \\ & & & & 0 & 0 & 0 & 0 & I & & * \\ & & & & & & & & 0 & \ddots & \vdots \\ & & & & & & & & & \ddots & I \\ & & & & & & & & & & 0 \end{bmatrix}, \qquad (3.37)$$

$$\begin{bmatrix} I & & & & & & \\ & I & & & & & \\ & & \ddots & & & & \\ & & & I & & & \\ & & & & \boxed{\begin{matrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{matrix}} & & \\ & & & & & I & \\ & & & & & & \ddots \quad I \end{bmatrix}. \qquad (3.38)$$

*The enclosed system is a solvable semiexplicit index-*1 *system* ($\bar{A}_{22}$ *invertible*) *with state-space dimension* $\sum_{i=1}^{r-1}(n_i - n_r)$; *thus the solution manifold for* (1.1) *has the same dimension. If r is odd* (*even*) *then* $q = p$ ($q = p+1$), *while if* $r = 2$, *the circled block* $\tilde{I}$ *is not present. All circled blocks in* (3.37), (3.38) *are* $n_r \times n_r$.

Thus Theorem 3.4 is independent confirmation of Proposition 3.2 and the fact that (3.25) and (3.26) specify the total solution manifold for H_r systems. However, (3.37) and (3.38) do not involve derivatives of the constraint, while (3.25), (3.26) do. Also note that (3.37), (3.38) can be written as

$$\begin{bmatrix} N_1 & G_1 & G_2 \\ 0 & E(t) & G_3 \\ 0 & 0 & N_2 \end{bmatrix} z' = \begin{bmatrix} I & 0 & 0 \\ 0 & F(t) & 0 \\ 0 & 0 & I \end{bmatrix} z + g, \tag{3.39}$$

where N_1, N_2 are structurally nilpotent and (E, F) is smoothly index 1. Hence H_r systems can be interpreted as a semiexplicit g-integrator [the dynamics of which determine the dynamics of (1.1), (3.1)] embedded between two g-differentiators of the transformed input $Z(t)f(t)$.

With the exception of the block diagonal time varying coordinate changes, all transformations in the reduction of H_r to (3.39) are BDF invariant. We strongly suspect the BDF invariance of these coordinate changes for H_r systems, but until this can be established the convergence behavior of BDF must be analyzed directly, rather than by appealing to the form (3.39) [3, 15, 27].

4. COMMENTS

In the literature, other Hessenberg-like structural forms have been studied in a control-theory context [16, 37]. However, these forms generally involve either the controllability or observability pairs (A, B) or (A, C) respectively, whereas the form defined in this paper is in terms of the realization triple (A, B, C). From Example 3.2, Hessenberg triples are neither controllable nor observable in general, but can be either in some special cases. Our form can also be distinguished by the invertibility condition on Π_r.

Finally, the concepts of g-integrator and g-differentiator have been introduced in this paper in a linear context. Extending the concept of g-integrator to nonlinear systems is quite easy. We say the *fully implicit* DAE

$$F(x', x, t) = 0 \tag{4.1}$$

is a nonlinear g-integrator if $\partial F/\partial x'$ is nonsingular or the pair $(\partial F/\partial x', \partial F/\partial x)$ evaluated along trajectories $(x'(t), x(t), t)$ is smoothly index 1. Since the input is not clearly identifiable in (4.1), we make no reference to it in the definition. Such systems are theoretically solvable, and the application of BDF produces convergent solutions [24]. However, extending the concept of g-differentiator poses some interesting problems. First, just as in the case for the convergence of BDF, necessary and sufficient conditions for the theoretical solvability of higher-index systems are not known, and thus cannot provide a basis for the definition. One could define a g-differentiator as in (2.2) where the components of N contain nonlinear terms of the form $N_{ij} = N_{ij}(x_1, x_2, \dots x_{i-1}, t)$. This imposes a somewhat linear structure on the system which does not in general exist for nonlinear DAEs. Furthermore, for more general nonlinear systems whose linearizations have embedded linear g-differentiators (e.g., the nonlinear Hessenberg form), it is not clear which transformations (if they exist) produce the chain of g-differentiators and g-integrators as in (3.39). For example, consider the nonlinear H_2 system

$$
\begin{aligned}
x' &= f(x, u, t), \\
0 &= g(x, t), \\
&\frac{\partial g}{\partial x}\frac{\partial f}{\partial u} \text{ invertible.}
\end{aligned}
\tag{4.2}
$$

If $n = \dim(x) = \dim(g)$, then $\dim(u) = n$ and we have a situation which is analogous to that in (3.15). That is, the implicit-function theorem can be applied to reduce (4.2) *implicitly* to

$$
\begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix}\begin{bmatrix} x' \\ u' \end{bmatrix} + \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} v(t) \\ w(v'(t), v(t), t) + v'(t) \end{bmatrix}. \tag{4.3}
$$

Even if only implicitly defined, it is clear which variables are differentiated and precisely how the differentiation occurs. On the other hand, such a simple reduction is not immediately available when $\dim(x) \neq \dim(g)$, and it may not exist in terms of linear-algebraic manipulations (e.g., row operations) or even nonlinear coordinate changes.

The author thanks the referees for their careful reading of the original manuscript.

REFERENCES

1 M. Athans and P. Falb, *Optimal Control: An Introduction to the Theory and Its Applications*, McGraw-Hill, New York, 1966.

2 D. J. Bell and D. H. Jacobson, *Singular Control Problems*, Academic, New York, 1975.

3 K. E. Brenan, Numerical simulation of trajectory prescribed path control problems, *IEEE Trans. Automat. Control* AC-31(3):266–269 (Mar. 1986).

4 K. E. Brenan and B. E. Engquist, Backward Differentiation Approximations of Nonlinear Differential-Algebraic Systems, Aerospace Corp. Report ATR-85(9990)-5, 1985.

5 S. L. Campbell, One canonical form for higher index linear time varying singular systems, *Circuits Systems Signal Process.* 2(3):311–326 (1983).

6 S. L. Campbell, Regularizations of linear time varying singular systems of differential equations, *Automatica* 20(3):365–370 (1984).

7 S. L. Campbell, *Singular Systems of Differential Equations I, II*, Pitman, Marshfield, Mass., 1982.

8 S. L. Campbell and K. Clark, Order and the index of singular time-invariant linear systems, *Systems Control Lett.* 1(2):119–122 (1981).

9 S. L. Campbell and K. Clark, Singular control problem structure and the convergence of backward differentiation methods, in *Proceedings of the International Conference on Mathematical Theory of Networks and Systems*, Phoenix, 1987, to appear.

10 K. Clark, Difference methods for the numerical solution of time varying singular systems of differential equations, *SIAM J. Algebraic Discrete Methods* 7(2):236–246 (1986).

11 K. Clark, Difference methods for the numerical solution of singular systems of differential equations, II: Dual properties of BDF and modified BDF with some applications, in preparation.

12 K. Clark, The Numerical Solution of Linear Time Varying Singular Systems of Differential Equations, Ph.D. Dissertation, North Carolina State Univ., Raleigh, 1986.

13 K. Clark, The numerical solution of some higher index time varying semistate systems by difference methods, *Circuits Systems Signal Process.* 6(1):261–275 (1987).

14 K. Clark, The numerical solution of index one singular systems by backward differentiation methods: Asymptotic expansions for the discretization error with applications to higher index systems, in preparation.

15 K. Clark, A structural form for higher index semistate equations, II: Convergence of backward differentiation formulas for the Hessenberg form, in preparation.

16 D. J. Clements and B. D. O. Anderson, *Singular Optimal Control: The Linear*

Quadratic Regulator Problem, Lecture Notes in Control and Information Sciences, No. 5, Springer, New York, 1978.

17 V. Dolezal, The existence of a continuous basis of a certain linear subspace of E_r which depends on a parameter, *Casopis, Pest. Mat.* 89 (1964).

18 I. Duff and C. W. Gear, Computing the structural index, *SIAM J. Algebraic Discrete Methods* 7(4):594–603 (1986).

19 B. A. Francis, The optimal linear-quadratic time-invariant regulator problem with cheap control, *IEEE Trans. Automat. Control* AC-24:616–621 (1979).

20 F. Gantmacher, *Theory of Matrices*, Chelsea, New York, 1960.

21 C. W. Gear, Simultaneous numerical solution of differential-algebraic equations, *IEEE Trans. Circuit Theory* CT-18(1):89–95 (Jan. 1971).

22 C. W. Gear, G. K, Gupta, and B. Leimkuhler, Automatic integration of Euler-Lagrange equations with constraints, *J. Comput. Appl. Math.* 12&13:77–90 (1985).

23 C. W. Gear and L. R. Petzold, Differential/algebraic systems and matrix pencils, in *Matrix Pencils*, Lecture Notes in Math., Vol. 973, Springer, New York pp. 75–89.

24 C. W. Gear and L. R. Petzold, ODE methods for the solution of differential/algebraic systems, *SIAM J. Numer. Anal.* 21(4):716–728 (1984).

25 D. H. Jacobson, Totally singular quadratic minimization problems, *IEEE Trans. Automat. Control* AC-16(6):651–657 (Dec. 1971).

26 A. J. Laub and A. Linnemann, Hessenberg forms in linear systems theory, in *Computational and Combinatorial Methods in Systems Theory* (C. Byrnes and A. Lindquist, Eds.), Elsevier Science (North Holland), Amsterdam, 1986, pp. 229–244.

27 P. Lotstedt and L. R. Petzold, Numerical solution of nonlinear differential equations with algebraic constraints: Convergence results for the backward differentiation formulas, *Math Comp.* 46(174):491–516 (1986).

28 N. Harris McClamroch, Singular Systems of Differential Equations as Dynamic Models for Constrained Robot Systems, Tech. Report RSD-TR-2-86, Center for Research on Integrated Manufacturing, College of Engineering, Univ. of Michigan, Ann Arbor, 1986.

29 R. W. Newcomb, The semistate description of nonlinear time variable circuits, *IEEE Trans. Circuits and Systems* CAS-28:62–71 (Jan. 1981).

30 J. F. Painter, Solving the Navier-Stoke Equations Using LSODI and the Method of Lines, Report UCID-19262, Lawrence Livermore National Lab., Livermore, Calif., 1981.

31 L. R. Petzold, Differential-algebraic equations are not ODEs, *SIAM J. Sci. Statist. Comput.* 3(3):367–384 (1982).

32 L. R. Petzold and P. Lotstedt, Numerical solution of nonlinear differential equations with algebraic constraints, II: Practical implications, *SIAM J. Sci. Statist. Comput.* 7(3):720–733 (1986).

33 J. Roberge, *Operational Amplifiers, Theory and Practice*, Wiley, New York, 1975.

34 L. M. Silverman, Inversion of multivariate linear systems, *IEEE Trans. Automat. Control* AC-14(3):270–276 (June 1969).
35 R. F. Sincovec, B. Dembart, M. A. Epton, A. M. Erisman, P. Lu, J. W. Manke, and E. L. Yip, Solvability of Large Scale Descriptor Systems, Final Report, Boeing Computer Services Co., Tukwila, Wash., June 1979.
36 I. Stakgold, *Green's Functions and Boundary Value Problems*, Wiley-Interscience, New York, 1979.
37 C. F. Van Loan, Using the Hessenberg decomposition in control theory, in *Algorithms and Theory in Filtering and Control* (D. C. Sorensen and R. J. Wets, Eds.), Math. Programming Stud. No. 18, North-Holland, Amsterdam, pp. 102–111.

Received 3 February 1987; final manuscript accepted 7 June 1987